ProCite in Libraries

Applications in Bibliographic Database Management

Edited by Deb Reneé Biggs



Learned Information, Inc. Medford, NJ 1995

- Division, American Library Association, 1991), 59-83.—See in particular section headed Summary of Research on Subject Access, 68-70.
- 3. See, among others: Carol Mandel, "Library Catalogs in the 21st Century," ARL: a Bimonthly Newsletter (September 9, 1992):1-4, and Marcia J. Bates, "Subject Access in Online Catalogs: A Design Model," Journal of the American Society for Information Science 37, no. 6 (1986):357-376.
- For a good concise explanation of methods for finding material on women see Susan Searing, *Introduction to Library Research in Women's Studies*, (Boulder, CO: Westview, 1985).

Chapte

Creating an Interdisciplinary Mental Health Database

Fred Lerner

National Center for Post-Traumatic Stress Disorder, White River Junction, Vermont

ABSTRACT

No single bibliography, printed or computerized, covers the literature on traumatic stress in its entirety. When the National Center for Post-Traumatic Stress Disorder was established in 1989, one of its first projects was the creation of a comprehensive database on the subject. ProCite was chosen as the software with which to implement this first of its kind database which would eventually be available worldwide as a BRS database (CHID) and for a time via the Internet.

Project Background

Post-traumatic stress disorder is a relatively recent addition to the nomenclature of psychiatry, but traumatic stress has been discussed in the mental health literature for over 100 years. Since the Vietnam war, the traumatic stress literature has increased enormously. It deals not only with the sequelae of combat but also with the consequences of natural and technological disaster and of crimes such as rape, incest, and torture. Both its producers and its users come from a wide range of disciplines: psychiatry, psychology, social work, criminology, law, religion, and many others.

No single bibliography, printed or computerized, covers the traumatic stress literature in its entirety. When the National Center for Post-Traumatic Stress Disorder (PTSD) was established in 1989, one of its first projects was the creation of a database to do this. In creating the PILOTS database, almost every decision was affected by the interdisciplinary nature of both the literature itself and the constituency to be served. This chapter describes some of the problems that were faced and the solutions adopted.

Selecting Software and Sources

The author's previous experience with bibliographic databases included several years as a user of files on DIALOG, Orbit, Reuters Textline, and STN, as well as the Dartmouth College Library Online System. As the tech-

nical information specialist in a small research and development firm, cataloging the company's technical library was also part of the job description. The process of automating the catalog was begun by using a primitive UNIX-based word-processing program to generate catalog cards on a line printer. When an MS/DOS-based microcomputer was made available, INMAGIC was used to create a bibliographic database. The original idea was to use the INMAGIC *Report* feature to print out a library catalog in book form, as company management objected to a catalog that could be searched only online. But when engineers working after hours found that they were able to use INMAGIC to find what they wanted through a relatively simple trial-and-error process—even when the information they had was seriously incomplete—they became enthusiastic about the idea of an online database.

Because of its affiliation with the Dartmouth Medical School, the National Center for PTSD chose to use Macintosh computers. As INMAGIC was not available for the Macintosh environment, a new, unfamiliar program had to be selected. We knew that we wanted one specifically designed for bibliographic applications, one that allowed for the multiple entries in some fields (such as author and descriptor) and the variable length of text in others that are characteristic of bibliographic descriptions. We had had sufficient experience with InfoStar and dBase III to know that a fixed-field database, no matter how powerful for payroll management or inventory control, was not well suited to bibliographic purposes.

Of the bibliographic database programs available for the Macintosh, ProCite¹ was selected for two reasons. First, we liked the appearance of the interface and the flexibility of the software. It was easy to learn, and using it did not require library experience; neither project staff nor users would need the extensive training demanded by many database programs. Second, the Project CORK Institute at Dartmouth College had been using ProCite to produce its database on clinical alcohol education. This provided sufficient proof that ProCite was robust enough to build and maintain a database of several thousand citations. It also offered the advantage of access to another local user who was two years further up the learning curve with the software than we would be.

The National Center for PTSD is the only unit in the entire Department of Veterans Affairs (VA) where a bibliographical database is produced. This offered advantages and disadvantages. The primary advantage was that there was no well-established VA way of doing things, which left us free to choose the best available software for the job and to configure it as we thought best. The primary disadvantage was that VA regulations and procedures did not contemplate the production of a bibliographical database. Funds for "automated data processing"—a term that includes within its purview not only computer hardware and software, but even database search services—were not available until near the end of the fiscal year.

This meant that one obvious method of keeping up with the traumatic stress literature—searching MEDLINE, PsycINFO, and other relevant databases—was not available. We had to find other ways of identifying material for possible inclusion in the database.

Current Contents was a logical choice for visually reviewing the literature of the field. We subscribed to three editions, all of which we received in paper: Clinical Medicine, Life Sciences, and Social and Behavioral Sciences. We made it a practice to scan these every week, page by page, using the back-of-the-issue indexes as a check to ensure that important papers were not inadvertently missed. Those articles that seemed relevant to our database were requested from their authors, using reprint request cards ordered from the Institute for Scientific Information, publisher of Current Contents.

We subscribed to *Current Contents on Diskette (CCOD)* when that became available, automating the printing of reprint requests. We took advantage of the program's compatibility with ProCite in two ways. First, by downloading author abstracts from *CCOD* we reduced our data entry effort. Second, we created a ProCite database of papers for which *CCOD* could not generate reprint requests or for which our requests had not been fulfilled. Sorting this database by journal facilitated obtaining those papers from libraries and document delivery services.

Another source was Dr. Arthur Arnold, a VA psychiatrist who had compiled a bibliography of publications on PTSD a few years before. Dr. Arnold was an enthusiastic supporter of our proposed database and sent us two large boxes of articles that he had collected in connection with his bibliography. This gave us immediate access to a substantial portion of the classic literature of traumatic stress.

Thesaurus Design

Now that we had documents to index and the software with which to produce and retrieve the descriptions, our database needed a name and an indexing vocabulary. At the same time that we were beginning our bibliographical work, another group within the National Center was planning a database of patient records, which would be used to support epidemiological research. To avoid confusion between the two projects, the name PILOTS, an acronym for "Published International Literature on Traumatic Stress," was chosen.

The indexing vocabulary was a thornier matter. Traumatic stress is an interdisciplinary field. Its literature is written by men and women trained in psychiatry, psychology, and social work, as well as less obvious fields such as art therapy, criminology, and public health. The users of that literature are equally varied in background. In addition to clinicians and researchers in the several mental health areas, potential PILOTS database users include policy

makers in public health, veterans' affairs, criminal justice, and foreign relations, PTSD patients, their families and friends, general practitioners in law, medicine, and social services who may encounter traumatic stress in their patients or clients, and students, teachers, and librarians for whom PTSD is a research topic rather than a presenting condition.

One possibility would have been to use an existing indexing vocabulary. The two obvious candidates were the Medical Subject Headings (MeSH) published by the National Library of Medicine and the Thesaurus of Psychological Index Terms published by the American Psychological Association (APA). The advantages of using one of these were that they were both well respected; both were maintained by a staff of experts, and both served a large community of searchers familiar with using them to locate papers of interest. But in both cases, the breadth of their coverage made them cumbersome to use in a relatively narrow branch of literature; and, while both had a term specific to PTSD ("stress disorders, post-traumatic" in MeSH; "post-traumatic stress disorder" in the APA thesaurus), neither appeared well suited to a detailed analysis of one narrow subdivision of an entire disciplinary domain. Another problem was that the adoption of one of these thesauri would suggest to many potential PILOTS users an alignment with one discipline. This did not seem like a good idea for a project that was essentially interdisciplinary.

The decision was ultimately made to produce a new thesaurus specifically for the PILOTS database. We read the standard books on thesaurus construction: Lancaster, Soergel, Aitchison and Gilchrist; studied several of the thesauri found at the Dartmouth library; and recalled personal experiences in using thesauri as an aid to formulating search strategies. From that, a good idea of what the thesaurus should look like began to take shape. The next step was to assemble the terms.

We began by looking through the *Journal of Traumatic Stress*. The official journal of the International Society for Traumatic Stress Studies, it is the only scientific journal devoted entirely to the field, and it is interdisciplinary in its contents. (The fact that it is not indexed by the National Library of Medicine was a selling point that would surface whenever someone questioned the need for the PILOTS database, given the existence of MEDLINE.) We included terms found in article titles and abstracts, and in the titles of papers cited in the journal. Extending our examination to other publications on traumatic stress increased the list. This process used up several packs of index cards and gave us several hundred terms that upon examination fell into a small number of distinct categories. There were stressors and people affected by them, effects of

their experiences and methods of treating those symptoms, diagnostic procedures, and policy issues.

These categories provided the basis for the hierarchical arrangement of concepts and relationships that is the basis of a thesaurus. To fill in gaps, the *Medical Subject Headings* and the *Thesaurus of Psychological Index Terms* were consulted to get an idea of what sort of terms might be needed and to use, whenever possible, terminology that had already been accepted by users and validated by experience. Thus, the terms for age groups in the PILOTS Thesaurus ("school age children," "preadolescents," etc.) are taken directly from the *Thesaurus of Psychological Index Terms*. (The APA advised that they had no objection to the limited use of its copyrighted publication.) For similar reasons the PILOTS terminology for psychiatric disorders is taken from the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-III-R*). In this case, *DSM-III-R* was considered a more authoritative source. A decision of this nature has to be made from time to time when working with an interdisciplinary literature.

Once a thesaurus is designed, it must be maintained. Fortunately, in recent years several microcomputer-based thesaurus maintenance programs have come on the market. One of the first of these was the *Liu-Palmer Thesaurus Construction System (TCS)*, which ensures that changes in the thesaurus are made to all cross-referenced terms and prints the thesaurus in a range of formats. We used it from the beginning and have been pleased with its simplicity and efficiency, and with the readiness of its producers to respond to the few problems we encountered. Our only regret is that it runs only on MS/DOS machines, requiring us to use translation programs like MacLinkPlus/PC to import its output into our Macintosh-based documentation files. In the best of all possible worlds, *TCS* would be integrated directly into ProCite's authority list feature, but there is nothing to indicate that this will ever happen.

Whenever the thesaurus is revised, we modify the Descriptors authority list in ProCite. We maintain supplementary authority lists to standardize the entries for specific incidents, organizations, and persons that we add to the Descriptors field. We also maintain an authority list of institutional affiliations (for the Author Affiliation field) and one of assessment instruments used in clinical or research reports (for the Instruments field).

Beginning the Database

At this point, six months into the project, we had a collection of documents to index, and we had a thesaurus of indexing terms. We were almost ready to begin entering material into the PILOTS database.

The predefined ProCite workforms were very well designed, but there were a couple of fields that we needed to add for our purposes. We wished

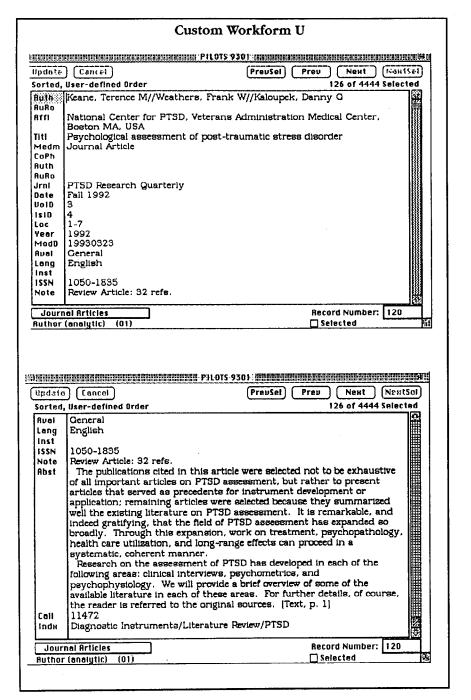


Figure 1

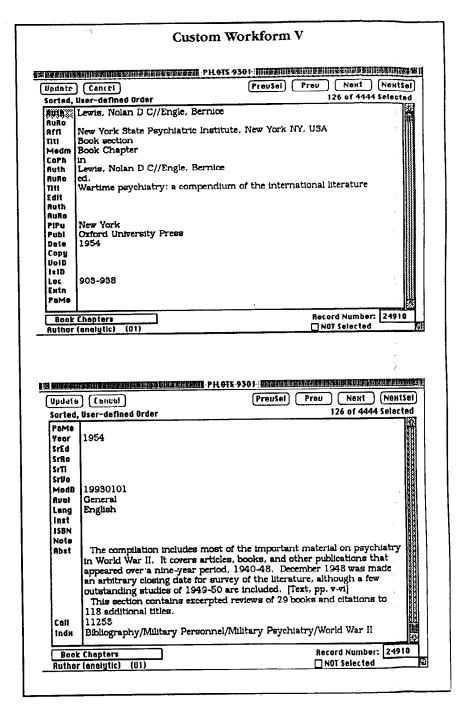


Figure 2

to indicate the language of each document and to indicate a source from which it might be available. (Some predefined ProCite workforms include an "Availability" field, but most do not.) Another field was needed to list the test and measurement instruments used in the work reported (a feature that is undoubtedly unique to the PILOTS database, and much appreciated by its users). To make updating the database easier, we wanted a field that would record the date on which each record was last modified. Two custom workforms were designed for this purpose: U (Journal Article) is based on C (Journal, Long) [FIGURE 1], and V (Book Chapter) is based on A (Book, Long) [FIGURE 2].

We plan to use the V workform for books, pamphlets, government documents, technical reports, and doctoral dissertations as well. Until we begin indexing audiovisual and computer resources, we will not need to add any more custom workforms.

One decision that was made at the beginning of the project was to avoid the use of abbreviations. The PILOTS database has always indexed papers under the authors' full names, using an authority list to standardize author entries. Instead of the user having to identify all possible forms under which an author's name might have appeared in print, we would undertake that responsibility as part of our preparation of the database. The goal—one more attainable in a small specialized bibliography than in a massive one like MEDLINE or PsycINFO—has been to expend our time and effort to minimize the user's. We give journal titles in full, even when the document in hand bears only an abbreviation. The only abbreviations we expect our users to deal with are two frequently-encountered ones that we use consistently in abstracts: PTSD (for post-traumatic stress disorder) and MMPI (for the Minnesota Multiphasic Personality Inventory).

We had purchased a scanner and optical character recognition (OCR) software in the hope that we could eliminate the need to type article abstracts into the database. This did not work out. Even the best microcomputer-based OCR software cannot deal effectively with the bold sansserif or italic typefaces that journal publishers like to use for setting abstracts; and in many cases we were working from third-generation photocopies, rather than pristine reprints or tearsheets. It did not take long for us to learn that a skilled and conscientious typist could do a much better and much faster job of data entry than could an Apple scanner and Caere's OmniPage. (Which is not to disparage that combination: we expect to use it extensively should we have the opportunity to produce a full-text product on CD-ROM.)

The preliminaries addressed, we were ready to enter citations into the database. The first paper indexed in PILOTS was "Toward rational pharmacotherapy for post traumatic stress disorder: an interim report," *American Journal of Psychiatry* 145(3): 281-285 (March 1988), by Matthew J, Friedman.

It seemed fitting to inaugurate the database with one of the most frequently cited papers in the PTSD literature.

Making the Database Accessible

A few hundred citations later, we had the beginnings of a database. But all our labor was of little use unless it was made available to others. How might we best do this?

This question was first addressed before the author was even hired. It was suggested that the database be mounted on a VA mainframe so that it could be accessible from any terminal or computer in the Veterans Administration. Those in charge of the VA mainframe offered to do this, if the National Center would pay for the estimated year of programmer time that this would require. That was the end of that.

The author's personal experience as a database searcher indicated that in order for PILOTS to be used by anyone but full-time PTSD researchers, the database should be mounted on an established host, such as BRS or DIALOG. The hazards of going through the paperwork involved in establishing an account with a new database vendor and in climbing the learning curve required to become even minimally proficient with a new command language cannot, however, be underestimated. If these were onerous in connection with ORBIT or STN, each of which offered access to scores of useful files, how much less inviting would it be to go through it all just to gain access to a small, specialized database? Nonetheless, we determined to find PILOTS a home on an established host with a large user base and a full-time help desk.

Subsequently, the National Center for PTSD joined the National Federation of Abstracting and Information Services (NFAIS), through whose publications and conferences we have learned many useful things about the technical, economic, and legal aspects of database production and distribution. At the first NFAIS conference that the author attended, contacts were made with representatives of databank hosts. DIALOG told us that the PILOTS database was too small and too specialized to fit onto their system; a similar response was elicited from BRS. We later found out during a conversation with a BRS vice president that, as a health-related database produced by a federal government health agency, we might find a place on BRS as part of CHID, the Combined Health Information Database.

CHID is a collection of subfiles, most of which are produced by various components of the National Institutes of Health (NIII) as part of its educational mission. Each subfile is devoted to a particular disease or group of diseases, and each emphasizes material for patients, family members, and the lay public, rather than the research literature indexed in MEDLINE.

Contact was made with the chairman of CHID and with its technical administrator, who encouraged our application for membership.

Establishing the PILOTS database on CHID was not easily accomplished. Most CHID files were produced at NIH—the technical production was performed by outside contractors under NIH supervision—and the CHID Board was used to thinking in terms of NIH goals and policies. There was some question as to how well PILOTS would fit in with other CHID files, especially as PILOTS was closer in form and content to MEDLINE or PsycINFO than to CHID. But after several months, and considerable communication, the CHID Board decided to encourage other federal health databases to join, and accepted PILOTS into membership.

Once again we were following in Project CORK's footsteps: its database was also on BRS, though as a separate file. We had to convert a ProCite database into a tape suitable for mounting on BRS, so we contracted with Project CORK, with whom we shared a Dartmouth affiliation, to help us through the process. CHID had its own format to which we were required to conform, and so the procedures which they devised for us differed slightly from Project CORK's. But the essentials were similar. This is the procedure that was developed:

- Working with a copy of the PILOTS database, use ProCite's search-and-replace command to convert any special characters (such as accents, umlauts, greater or less than signs, and percent signs) into standard characters.
- Do two searches, to identify records using each of our workforms.
- Use specially written punctuation files to turn each group into a text file incorporating CHID formatting requirements and eliminating PILOTS fields extraneous to CHID practice.
- Use Microsoft Word to complete the transformation into CHID format.
- With program written for us by Project CORK, use the Dartmouth College Time Sharing mainframe to check the text file for conformity to line length and other constraints, and place its contents on a magnetic tape.
- Send the completed tape, with a cover letter, to the CHID technical administrator in time to meet the deadline for quarterly updates.

Beginning with the second quarter, we had to repeat most of these steps in order to submit files representing records that had been modified since they were first sent to CHID. We searched on the "Modification Date" field that we had added to our custom workforms in order to identify these records.

In April 1991, eighteen months after work had begun on the project, the PILOTS database was publicly available for searching as the PTSD subfile of the Combined Health Information Database. There were 1,950 records in the file: only a fraction of the PTSD literature. We were careful to warn potential users of that, but still we were a valuable resource for locating literature on traumatic stress. We reported the availability of PILOTS in our PTSD Research Quarterly, we announced it to VA medical librarians and PTSD specialists through internal electronic mail groups, and we sent press releases to several newsletters and newspapers serving psychiatrists, psychologists, and medical librarians.

Documentation and Database Distribution

A database is not as useful without documentation, so we prepared a *PILOTS User's Guide*. Designed for librarians and expert searchers, it presumed a familiarity with database searching and an understanding of BRS or ProCite search commands. Its major components were a field-by-field discussion of the contents of PILOTS records, and the PILOTS Thesaurus presented in both hierarchical and alphabetic forms, the latter including an abundance of cross-references. We sent a copy of the *Guide* to each VA medical library, and announced that others could purchase it from the National Technical Information Service (NTIS). Once again our intention was to divert as much support activity as possible from us to agencies with adequate resources to do the job properly: we had neither the capability nor the desire to be booksellers, and the arrangement with NTIS was self-financing.

Another arrangement with NTIS was intended to provide an alternative to pay-by-the-minute searching via BRS. We announced that the database would be available for purchase on floppy disk. Our intention was to provide NTIS with a new copy of the database and its associated index files on a quarterly basis, in both Macintosh and MS-DOS formats. They would duplicate these whenever they received an order. We estimated that in order to recover costs it would be necessary to charge \$60 for a copy of the database or \$200 for an annual subscription. In addition, the purchaser would need to have a copy of ProCite or of some other database program that could import ProCite records.

So far as we could tell, use of the database was very modest. Though we knew the extent to which CHID was searched, there was no way of determining how much use any particular subfile received. We received no figures from NTIS about database sales; to date, none had been sold. It took several months to iron out technical aspects of getting the database to NTIS and reproducing it for customers—the file was much too large to

fit on a single disk, and the use of segmenting or compression programs brought up issues of licensing. This overwhelming lack of interest soon led us to withdraw the database from sale through NTIS. (By this time we had received the results of a survey conducted for us by students at Dartmouth's Tuck School of Business. This revealed that clinicians and researchers would prefer online access to PILOTS on an as-needed basis rather than purchasing the database on diskette.)

Instead, through the cooperation of the Kiewit Computer Center at Dartmouth College, the database was made available for downloading through anonymous file transfer protocol from one of Dartmouth's computers. PILOTS users with Internet access and expertise were invited to help themselves, but we never heard from anyone who had bothered to do so. We recently discontinued this availability, in order to maintain the integrity of the database and to eliminate one potential obstacle to licensing other organizations' abstracting and indexing for use in our database.

The PILOTS database was now available to users worldwide. We had produced and distributed a *User's Guide* which would allow a trained searcher to find citations in the database. And we were continuing to identify, acquire, and index additional papers on traumatic stress. By July 1992 we had over 4,000 records in the database: this was still not a complete representation of the literature, but certainly the largest single bibliography on traumatic stress available. The next goal was to make it more available.

We were under no mandate to earn money or even to recover costs from database users. Our success would be measured in terms of how much use PILOTS received, especially within the Department of Veterans Affairs. One way of increasing usage would be to make it available for free; and the simplest way of doing that would be to add the PILOTS database to the files available for searching on the Dartmouth College Library Online System (DCLOS). Dartmouth had long been a pioneer in academic computing and in library automation. Because of our affiliation with the Dartmouth Medical School, we were able to place PILOTS on DCLOS, using essentially the same procedures that had been developed for CHID. (DCLOS uses BRS Search as its command language.) After a brief experimental period we announced its availability at the annual conference of the International Society for Traumatic Stress Studies in October 1992. A four-page "PILOTS at Dartmouth" supplement to the Fall 1992 PTSD Research Quarterly brought the word to 4,500 readers worldwide, who were already aware of the database through a "PILOTS Update" column in each issue. Announcements were sent via print and computer mail to newsletters and electronic list servers in several relevant disciplines. As we expected, interest in the database has increased substantially, and usage has grown to the point of requiring regular provision of time for user support.

We have not, by any means, come to an end of our work. There are thousands more documents out there: papers we have identified through searching other databases and consulting other bibliographies, publications on "soldier's heart," "shell shock," "war neurosis," and other pre-DSM (Diagnostic and Statistical Manual of Mental Disorders) manifestations of traumatic stress; plus, there are materials in at least twenty languages other than English as well. We are working with groups in Europe with the goal of establishing common standards and cooperating to ensure bibliographical coverage of the world's traumatic stress literature. And we hope to induce sociologists and information scientists to use PILOTS as a tool in exploring interdisciplinary communication:

We have accomplished a fair piece in three years, but there is a lot left for us to do. In a world vulnerable to earthquake and hurricane, a world that has not abolished rape and incest, a world that can still produce Bosnias and Somalias, we can expect the study and treatment of post-traumatic stress disorder to continue for a long time. The PILOTS database offers an example of how limited resources and a bit of ingenuity can make a contribution to dealing with a major social problem.

Notes

1. We began our work using version 1.34, but converted to version 2.0 when it became available. We are now using version 2.1.

Bibliography

Aitchison, Jean and Gilchrist, Alan. *Thesaurus construction: a practical manual*, 2nd ed. London: ASLIB, 1987.

Lancaster, F.W. Vocabulary control for information retrieval, 2nd ed. Arlington, VA: Information Resources Press, 1986.

Soergel, Dagobert. Indexing languages and thesauri: construction and maintenance. Los Angeles: Melville, 1974.